

O papel da classificação na recuperação da informação: classificação facetada

Malvina Vianna Rosa

Biblioteca da Faculdade de Odontologia
Universidade Federal do Rio Grande do Sul
Porto Alegre, RS

Resumo — Entre as funções próprias das classificações a mais importante é a de possibilitar a construção de linguagens de recuperação, visando a obter informações armazenadas. Toma-se evidente a ineficiência das classificações tradicionais no que diz respeito à qualidade e estruturação das linguagens a que dão origem. As classificações facetadas, baseadas na análise de facetas, apresentam terminologia atual e mostram a realidade do pensamento dos especialistas sobre seus assuntos. Pela observância coerente de uma ordem de citação, proporcionam um controle valioso, além de determinar uma ordem de arquivamento útil ao leitor.

A descrição do assunto de um documento começa pela atribuição de palavras que representam o assunto e atuam como chaves na recuperação da informação do documento. Estas palavras-chave são uma condensação do texto, portanto, não traduzem toda a informação encontrada. Sendo guias parciais e imperfeitos, somente por aproximação conduzem ao texto.

Na recuperação da informação, as palavras que representam a consulta do leitor são cotejadas com as palavras-chave usadas pelo indexador para representar o documento. As palavras empregadas na formulação da consulta podem não coincidir com aquelas utilizadas na indexação, resultando na não recuperação da informação contida em textos existentes na coleção.

A maneira de minorar tal problema é aumentando o número de palavras que representam a consulta. Esta generalização é, via de regra, feita empiricamente, isto é, aquele que realiza a consulta criará alternativas. Usando a classificação, ou seja, se classificarmos as palavras-chave obtaremos alternativas e oportunidades de expansão mecânica da consulta. A classificação, quando empregada na recuperação da informação, também nos permite a realização do oposto. Se a consulta formulada for demasiadamente geral, recuperando grande número de documentos não desejados, a simples verificação do catálogo classificado nos conduzirá ao assunto específico procurado. A reunião de assuntos correlatos, que so-

mente a classificação possibilita, uma vez que tais assuntos ficam dispersos no catálogo alfabético de assuntos, constitui-se na maior vantagem do uso da classificação na recuperação da informação.

Os esquemas de classificação eram, tradicionalmente, desenvolvidos para suprir duas necessidades consideradas fundamentais:

- 1) a arrumação dos livros nas estantes, segundo seus relacionamentos lógicos; e
- 2) a ordenação sistemática das formas representativas dos documentos, em bibliografias e catálogos.

Como linguagens de recuperação entendemos os registros de informação organizados no intuito de uma fácil recuperação. Por isto, apresentam riqueza de termos genéricos e específicos. Os sistemas de classificação são vistos como instrumentos capazes de originar pelo menos três diferentes tipos de linguagens de recuperação, a saber:

- a) catálogo classificado mais índice alfabético complementar; b) catálogo alfabético de assuntos; c) thesaurus.

A compreensão da estrutura e funcionamento das linguagens de recuperação nos leva a descobrir a ineficiência dos esquemas de classificação que se fundamentam em princípios da lógica aristotélica, apresentando classes já compostas, porém limitadas, constituindo-se em listagens repetidas e enumerativas.

As informações contidas nos documentos devem ser organizadas tendo em vista sua melhor comunicação. O arranjo sistemático deve representar a estrutura da realidade descrita nos documentos, objetivando a arrumação da informação dentro de uma seqüência útil. A função da classificação na recuperação da informação é apresentar ao leitor um mapa ou esquema detalhado do assunto que lhe interessa.

As classificações facetadas, baseadas na análise de facetas, adotam valiosos princípios e técnicas de controle para a recuperação da informação. O termo análise de facetas foi introduzido, pela primeira vez em classificação bibliográfica, por Ranganathan, indicando a técnica de separar os vários elementos de um assunto complexo em relação a um conjunto de conceitos fundamentais abstratos. Para Ranganathan a análise de facetas é “o processo mental através do qual são enumerados os possíveis conjuntos de características que podem formar as bases da classificação de um assunto. Este processo mental também determina a medida exata em que os atributos concernentes aparecem no assunto” (4).

As causas principais que determinam a ineficiência dos atuais sistemas gerais de classificação e a conseqüente elaboração de sistemas especiais, são:

- a) a ausência de detalhe para a especificação de assuntos altamente complexos;

- b) embora apresentando uma cobertura ampla de assuntos, não fornecem os pontos de vista específicos da biblioteca ou serviço de informação;
- c) assuntos muito complexos requerem grande flexibilidade na combinação dos termos, o que é impossível com os sistemas existentes;
- d) quando flexíveis, usam notação demasiadamente longa e complicada;
- e) geralmente a ordem de arquivamento não é de utilidade para o usuário, pois se distancia da realidade (5).

Na construção de uma classificação facetada examina-se uma amostragem da literatura que está sendo produzida sobre o assunto (1). Posteriormente, analisam-se os termos, sendo os mesmos distribuídos em grupos homogêneos ou facetas. Chamamos facetas a coleção de termos ou vocábulos que apresentam o mesmo relacionamento com o assunto global, refletindo a aplicação de um princípio básico de divisão. A análise de facetas organiza os termos, permitindo construir classes por síntese em lugar de divisão. As palavras são relacionadas como elementares e não previamente combinadas, constituindo complexos para serem incluídos no esquema. Assim, na classificação facetada de Odontologia, encontramos termos tais como: boca, extração, cimentação, língua, incisão, inflamação, etc. Estas palavras foram agrupadas nas seguintes facetas:

Entidades: boca, língua, gengiva, dentes, caninos, molares, etc.

Processos: trauma, inflamação, fratura, tumor, neoplasia, etc.

Operações: construção, cimentação, restauração, incisão, etc.

As facetas obtidas são inerentes ao assunto. Dentro de cada faceta, os focos ou termos, que as constituem, são suscetíveis de novos agrupamentos pela aplicação de outras características divisionais, originando subfacetadas. As palavras ou termos dentro de uma subfaceta são mutuamente exclusivas, isto é, não se sobrepõem para a formação de assuntos compostos. Exemplifiquemos com termos como escovação, exodontia, dentaduras completas. Todos são identificados como pertencendo à faceta operação. Mas cada palavra citada requer a aplicação de outras características divisionais, originando subfacetadas como:

Operações (faceta)

Odontologia preventiva (subfaceta)

Escovação

Cirurgia (subfaceta)

Exodontia

Prótese dentária (subfaceta)

Dentaduras completas

A grande maioria dos assuntos procurados para recuperação são compostos.

O problema central da indexação é a predicabilidade na localização das

classes. Ela é alcançada pelo estabelecimento de uma ordem de citação e pela sua coerente observância. Para a Colon Classification foi formulada a ordem de citação conhecida pela sigla PMEST (Personalidade/Matéria/Energia/Espaço/Tempo). O Grupo de Pesquisas em Classificação, na Inglaterra, estabeleceu a ordem padrão de citação (2). Tal ordem, sendo mais detalhada, apresenta em seu contexto os artificios denominados por Ranganathan de “ciclos e níveis” (*rounds and levels*) (3). Contudo, a ordem de citação determinada para um assunto deve evidenciar a maneira como os especialistas no assunto pensam, isto é, a maneira como os assuntos serão procurados.

A ordem de arquivamento determina como os documentos serão arquivados, ou seja, a seqüência em que aparecerão as informações contidas nos documentos. Se tal ordem for inversa à ordem de citação, garantirá a ordenação do geral antes do específico, correspondendo, assim, à expectativa do grande público. Por último, queremos nos referir à notação. Sua principal função é mecanizar a seqüência das classes. As notações não hierárquicas, além de mais curtas, cumprem a função básica de mecanização da seqüência.

O campo dos conhecimentos desenvolveu-se em complexidade e os documentos especializados consistem mais na discussão sobre relações e interconexões existentes entre duas entidades do que da mera descrição das entidades. A análise de facetas, correspondendo ao modo como o especialista vê seu assunto, tem um papel vital a desempenhar nas bibliotecas e serviços de informação.

Abstract

The role of classification in information retrieval:
faceted classification

The most important function of a classification system is to permit the construction of retrieval languages which will lead the user to the information he wants. The inefficiency of traditional classification systems is stressed by the poor quality and the structure of the retrieval languages they originate. Faceted classifications make use of modern terminology and display the way the specialists actually think about their subjects. They provide a useful control by the consistent adherence to a given citation order and adopt a sequence of arrangement which is helpful to the user.

REFERÊNCIAS

1. FOSKETT, D.J. The construction of a faceted classification for a special subject. In:----- . *Science, humanism and libraries*. London, Crosby Lockwood, 1964, p. 143-165.
2. MILLS, J. Library classification. *Journal of Documentation* 26 (2) : 120-160, June 1970.
3. RANGANATHAN, S.R. *Colon classification: basic classification*. 6. ed. Bombay, Asia Publishing House, 1960, 422 p.
4. ----- . *Philosophy of library classification*. Copenhagen, Ejnar Munksgaard, 1951, p. 23. (Library research monographs, 2).
5. VICKERY, B.C. *Faceted classification: a guide to construction and use of special schemes*. London, Aslib, 1970. 70 p.